On a class of recursive games

Nicolas Vieille

ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

On a class of recursive games

Nicolas Vieille

Août 2001

Cahier n° 2002-005

LABORATOIRE D'ECONOMETRIE

1rue Descartes F-75005 Paris

(33) 1 55558215

http://ceco.polytechnique.fr/

mailto:labecox@poly.polytechnique.fr

On a class of recursive games

Nicolas Vieille[1]

Août 2001

Cahier n° 2002-005

Résumé: Ce texte est à paraître dans les Proceedings of the NATO ASI on Stochastic Games (volume préparé sous la direction de A. Neyman et S. Sorin). Il traite de l'existence de paiements d'équilibres dans une classe de jeux stochastiques récursifs.

Abstract: This is a contribution to appear in Proceedings of the NATO ASI on Stochastic Games (editors: A. Neyman and S. Sorin). It deals with the existence of equilibrium payoffs in a class of recursive stochastic games.

Mots clés : jeux stochastiques, jeux récursifs

Key Words : stochastic games, recursive games

Classification AMS: 91A05, 91A15

[1] Ecole Polytechnique and Département Finance et Economie, HEC, 78 351 Jouy-en-Josas, France.e-mail: vieille@hec.fr

On a class of recursive games

Nicolas Vieille

December 12, 2001

In this chapter, we deal with a class of two-player, recursive games. Recall that a recursive game is a stochastic game such that $r(z, a, b) = 0$ whenever z is not an absorbing state. The games we consider have in addition the following properties:

F.1 in any absorbing state, the payoff to player 2 is positive;

F.2 for every initial state, and profile (α, β) where β is fully mixed (*i.e.* $\beta_z(b) > 0$ for every $(z, b) \in S \times B$), the induced play reaches an absorbing state in finite time.

We present the proof of the following result.

Theorem 1 *Any such game has a uniform equilibrium payoff.*

The interest of this specific class of games lies in the fact that the problem of equilibrium payoff existence for general two-player games can be reduced to this class. This reduction was done in the previous chapter. One saw there that one could assume in addition absorbing payoffs of player 1 to be negative. It is not clear how to use this additional property.

Let Γ be such a recursive game. The basic idea of the proof is to construct a family $(\Gamma_\varepsilon)_{\varepsilon>0}$ of approximating games, in which player 2's strategy choice is restricted. For each game in the family, we define a modified best-reply map, and apply a fixed-point argument to derive a stationary profile $(\alpha^\varepsilon, \beta^\varepsilon)$. Moreover, $(\alpha^\varepsilon, \beta^\varepsilon)$ is a Puiseux profile (as a function of ε). The upshot is to prove that $\lim_\varepsilon \gamma(\alpha^\varepsilon, \beta^\varepsilon)$ is a uniform equilibrium payoff of Γ. We use extensively the tools introduced in a previous chapter by Solan [6].

The chapter is organized as follows. We start with an example that shows that a stationary ε-equilibrium needs not exist for such games. This contrasts with the case of zero-sum recursive games, where stationary ε-optimal strategies do exist (Everett [2]). In Section 2, we define the constrained games, and the modified best-reply map. The discussion there is complete, and is a simplification of the proof in Vieille [8][1]. By contrast, Section 3 contains essentially no complete proof. Our goal there is to give a detailed (though non-rigourous) discussion of a specific case of a game with two non-absorbing states. This discussion contains already all the features of the general proof, but the simplicity of the setup enables us to avoid many technicalities.

1 Example

Consider the game

$$z_1\ \begin{array}{|c|c|}\hline \swarrow & -2,1^* \\ \hline\end{array}$$

$$z_2\ \begin{array}{|c|}\hline \nearrow \\ \hline \frac{4}{5}\searrow + \frac{1}{5}(-3,3)^* \\ \hline\end{array}$$

$$z_3\ \begin{array}{|c|c|}\hline \nwarrow & -1,2^* \\ \hline\end{array}$$

It is a variation on a example due to Flesch, Thuijsman and Vrieze [3]. There are three non-absorbing states z_1, z_2, z_3. The game is a game of perfect information : in each state, only one player has to play. In states $z_1$and z_3 player 2 can either choose to go to z_2 (by playing the left column), or to go to an absorbing state (right column) with the indicated payoff. In state z_2, player 1 can either choose to go to z_2 (by playing the top row), or to play the bottom row, which results in a non-deterministic transition: with probability $\frac{4}{5}$, the play moves to z_3; it otherwise moves to an absorbing state with payoff vector $(-3,3)$. Payoff in the non-absorbing states is zero.

It is clear that this game has the payoff and transition features we assume in this lecture. For ε small enough, there is no profile (α,β) that is an ε-equilibrium for any initial state. Indeed, if α_{z_2} puts a positive probability on the bottom row, the unique $(\varepsilon-)$best reply of player 2 is the stationary strategy which chooses the left column in both states z_1 and z_3. The unique $(\varepsilon-)$best reply of player 1 to this strategy is the stationary strategy which chooses the top row in state z_2.

If α_{z_2}puts probability one on the top row, any $(\varepsilon-)$best reply of player 2 when the initial state is z_3 chooses the right column in state z_3. Given any such strategy, the unique $(\varepsilon-)$best reply of player 1 when the initial state is z_2 is the stationary strategy which chooses the bottom row in state z_2.

[1] I wish to thank Eilon Solan for very helpful suggestions.

This prompts the following question. Given z, is there a stationary ε-equilibrium for the initial state z ? This is the case in the example above. Whether this holds or not in general is an open problem.

2 Constrained games

We denote by Σ_s, T_s the sets of stationary strategies of the two players, and set

$$T_s(\varepsilon) = \{\beta \in T_s \text{ such that } \beta_z(b) \geq \varepsilon \text{ for every } z \in S^*, b \in B\}$$

Choose integers $n_0, ..., n_{|B|\times|S^*|}$ such that $n_0 = 0$, and $n_{p+1} > |S|(n_p + 1)$, for each p, and set $N = n_{|B|\times|S^*|}$ For each ε, we define a set-valued map on the convex compact set $\Sigma_s \times T_s(\varepsilon^N)$. Equivalently, we define a game Γ_ε in which player 2 is restricted to stationary strategies in $T_s(\varepsilon^N)$. The pleasant feature of this restriction is the following. For every pair $(\alpha, \beta) \in \Sigma_s \times T_s(\varepsilon^N)$, the induced play is absorbing. Therefore, the function $\gamma(z, \cdot, \cdot)$ defined by $\gamma(z, \alpha, \beta) = \lim_n \gamma_n(z, \alpha, \beta)$ is continuous over $\Sigma_s \times T_s(\varepsilon^N)$. The idea of restricting strategy spaces in order to recover continuity of the (limit) payoff function is not new. It was, for instance, used in Vieille [7] to study absorbing recursive games with properties **F1** and **F2**, and in Flesch, Thuijsman and Vrieze [3] to study general absorbing recursive games.

A natural idea is to look for a stationary equilibrium $(\alpha^\varepsilon, \beta^\varepsilon)$ in the game Γ_ε. The existence of such an equilibrium follows from standard arguments. One then investigates the asymptotic properties of $(\alpha^\varepsilon, \beta^\varepsilon)$. Are of particular interest the limit profile $(\alpha, \beta) = \lim_\varepsilon (\alpha^\varepsilon, \beta^\varepsilon)$, and the limit payoff $\gamma = \lim_\varepsilon \gamma(\alpha^\varepsilon, \beta^\varepsilon)$ (both exist up to a subsequence). One might hope to be able to construct an ε-equilibrium in the original game, by perturbing the limit profile (α, β) in an appropriate way. This is the approach followed by Solan [5]. It succeeds for games with at most *two* non-absorbing states. The drawback is that the equilibrium payoff one obtains differs from γ. This precludes any extension.

2.1 The modified best-replies

We define a product set-valued map $\Phi(\alpha, \beta) = \Phi_1(\beta) \times \Phi_2^\varepsilon(\alpha, \beta)$ on $\Sigma_s \times T_s(\varepsilon^N)$. Observe that for every (α, β) and every initial state $\overline{z}$, the probability of ending in the absorbing state z^* is a rational function of the variables

$\alpha_z(a), \beta_z(b),\ \ (z,a,b) \in S^* \times A \times B$. Therefore, $\gamma(\overline{z}, \alpha, \beta)$ is also a rational function of the same variables.

2.1.1 Definition of Φ_1

Define $\Phi_1(\beta)$ as the set of stationary best-replies of player 1 to β:

$$\Phi_1(\beta) = \{\alpha^*, \text{ such that } \gamma^1(z, \alpha^*, \beta) \geq \gamma^1(z, \alpha, \beta) \text{ for every } z \in S^*, \alpha \in \Sigma_s\}$$

The existence of such a best-reply is due to Blackwell [1].

Set $\gamma^1_M(z,\beta) = \gamma^1(z, \alpha^*, \beta)$, where α^* is any stationary best reply of player 1. Since any profile in $\Sigma_s \times T_s(\varepsilon^N)$ is absorbing, an element of $\Phi_1(\beta)$ is characterized by the property:

$$E\left[\gamma^1_M(\beta) | z, \alpha_z, \beta_z\right] = \max_A E\left[\gamma^1_M(\beta) | z, \cdot, \beta_z\right], \text{ for every } z \in S^*. \tag{1}$$

It follows that $\Phi_1(\beta)$ is a face of the polytope of stationary strategies of player 1. It is clear that Φ_1 is upper hemicontinuous, and has non-empty values.

2.1.2 Definition of Φ_2^ε

We now define $\Phi_2^\varepsilon(\alpha, \beta)$. Any action of any strategy in $\Phi_2^\varepsilon(\alpha, \beta)$ is to have a positive probability. We shall define a measure of the quality of a given action, and require that actions of different qualities have probabilities of different orders of magnitude. A natural candidate for measuring the quality of an action b in state z is the quantity

$$E\left[\gamma^2(\alpha, \beta) | z, \alpha_z, b\right]$$

it measures the expected payoff of player 2, when the initial state is z, and he plays b then β against α. This defines a good measure of how well actions perform in state z against α. However, when it comes to comparing actions across states z and z', it is unsatisfactory since it gets intertwined with the comparison of the two payoffs $\gamma^2(z, \alpha, \beta)$ and $\gamma^2(z', \alpha, \beta)$.

To disentangle the two comparisons, we define the *cost* of action b in state z against α by

$$c(b; z, \alpha, \beta) = \max_B E\left[\gamma^2(\alpha, \beta) | z, \alpha_z, \cdot\right] - E\left[\gamma^2(\alpha, \beta) | z, \alpha_z, b\right].$$

The following properties clearly hold:

P.1 $c(b; z, \alpha, \beta) \geq 0$ for every z, b, α, β; moreover, $\min_B c(\cdot; z, \alpha, \beta) = 0$;

P.2 For every b and z, the function $c(b; z, \cdot, \cdot)$ is semialgebraic (see the chapter by Neyman [4]).

Given (α, β), we denote by $C_0(\alpha, \beta), ..., C_L(\alpha, \beta)$ the partition of $S^* \times B$ into level sets for the function $c(\cdot; \cdot, \alpha, \beta)$, ranked by increasing cost (of course, the number $L + 1$ of level sets depends on (α, β)). Define $p_0 = 0$, and $p_l = \sum_0^{l-1} |C_i(\alpha, \beta)|$, for $0 < l \leq L$. Thus, for $(z_0, b_0) \in C_l(\alpha, \beta)$, p_l is the number of state-action pairs (z, b) that are strictly better than (z_0, b_0), *i.e.* $c(b_0; z_0, \alpha, \beta) > c(b; z, \alpha; \beta)$.

Define $\Phi_2^\varepsilon(\alpha, \beta)$ as the set of $\tilde{\beta} \in T_s(\varepsilon^N)$ such that for every $(z, b) \in C_l(\alpha, \beta)$, $0 \leq l \leq L$, one has

$$\varepsilon^{n^{p_{l+1}-1}} \leq \tilde{\beta}_z(b) \leq \varepsilon^{n^{p_l}}$$

By **P.1**, for every $z \in S^*$, there is $b \in B$, such that $(z, b) \in C_0(\alpha, \beta)$. It easily follows that $\Phi_2^\varepsilon(\alpha, \beta)$ is non-empty. The exact definition of $\Phi_2^\varepsilon(\alpha, \beta)$ is tailored for an application of Kakutani's theorem, and for getting semialgebraicity properties. What is truly important for the asymptotic analysis that we present later is the observation below, which follows immediately from the definition of $\Phi_2^\varepsilon(\alpha, \beta)$: for every $z, z' \in S^*$, $b, b' \in B$, and $\tilde{\beta} \in \Phi_2^\varepsilon(\alpha, \beta)$

$$c(b'; z', \alpha, \beta) > c(b; z, \alpha, \beta) \Rightarrow \tilde{\beta}_{z'}(b') \leq \varepsilon \left[\tilde{\beta}_z(b)\right]^{|S^*|} \tag{2}$$

2.1.3 Existence of a fixed point

By Kakutani's theorem, the product map $\Phi_1(\alpha) \times \Phi_2^\varepsilon(\alpha, \beta)$ has a fixed point in $\Sigma_s \times T_s(\varepsilon^N)$. Denote by

$$F(\varepsilon) = \{(\alpha, \beta) \in \Sigma_s \times T_s(\varepsilon^N), \alpha \in \Phi_1(\beta), \beta \in \Phi_2^\varepsilon(\alpha, \beta)\}$$

the set of these fixed points. By **P.2,** and the definitions of Φ_1 and Φ_2^ε, the set-valued map $\varepsilon \mapsto F(\varepsilon)$ is semialgebraic. Therefore, (see [4]) there exists a semialgebraic selection of F: for each $\varepsilon > 0$ small enough, there exists $(\alpha^\varepsilon, \beta^\varepsilon) \in F(\varepsilon)$, such that $\alpha_z^\varepsilon(a), \beta_z^\varepsilon(b)$ have Puiseux expansions in ε, for every $(z, a, b) \in S^* \times A \times B$.

3 Asymptotic analysis

3.1 General comments

We here consider the Puiseux profile $(\alpha^\varepsilon, \beta^\varepsilon)$ that was obtained in the previous section. Our chief goal is to prove that $\gamma = \lim_\varepsilon \gamma(\alpha^\varepsilon, \beta^\varepsilon)$ is an equilibrium payoff of the game. We set $(\alpha^0, \beta^0) = \lim_\varepsilon (\alpha^\varepsilon, \beta^\varepsilon)$.

Recall from [6] that the map $\varepsilon \mapsto (\alpha^\varepsilon, \beta^\varepsilon)$ induces a hierarchical decomposition of S^* into (a forest of) communicating sets, which reflects how the behavior of the Markov chain induced by $(\alpha^\varepsilon, \beta^\varepsilon)$ depends on ε.

A communicating set C is defined by the property that, given an initial state in C, the probability under $(\alpha^\varepsilon, \beta^\varepsilon)$ that the play will reach any state in C before it leaves C goes to one as ε goes to zero. The leaves of the forest (*i.e.*, the smallest communicating sets) coincide with the subsets of S^*, which are ergodic w.r.t. (α^0, β^0). The roots are the largest communicating sets.

We denote by $D^1, ..., D^H$ the roots of the forest. We set $T = S^* \backslash (D^1 ... \cup D^H)$: T is the set of states which belong to no communicating set.

For $1 \leq h \leq H$, we denote by Q^h the distribution of exit from D^h, as defined by $(\alpha^\varepsilon, \beta^\varepsilon)_\varepsilon$: $Q^h(z)$ is the limit (as ε goes to zero) of the probability under $(\alpha^\varepsilon, \beta^\varepsilon)$ that, starting from D^h, z is the first state outside D^h that is reached.

Consider the Markov chain over the state space $\{\{D^1\}, ..., \{D^H\}\} \cup T \cup A$ whose transition function $\mathbf{p}$ is Q^h in $\{D^h\}$, and $p(\cdot|z, \alpha^0, \beta^0)$ for $z \in T$.

Lemma 2 *The Markov chain with transition function $\mathbf{p}$ is absorbing.*

Proof. there would otherwise be a communicating set included in T, or a communicating set which strictly contains some D^h. In either case, this would contradict the fact that $D^1, ..., D^H$ are the roots of the forest. ■

The next proposition presents no difficulty. It uses the previous lemma.

Proposition 3 *Assume that: (1) for each $z \in T$, the pair of mixed actions (α^0_z, β^0_z) is a Nash equilibrium of the matrix game with payoff $E[\gamma|z, \cdot, \cdot]$;(2) for each $1 \leq h \leq H$, the distribution Q^h is a controllable exit distribution for γ. Then γ is an equilibrium payoff.*

Proof. let us briefly describe a profile (σ, τ) which supports γ. Whenever the current state belongs to T, the profile (σ, τ) plays like (α^0, β^0) (irrespective of past play). Whenever the game enters some set D^h, the players switch to the profile (σ^h, τ^h) associated with the controllable exit distribution Q^h. Finally, the players switch to punishment strategies if the game has not entered an absorbing state by stage N_0, where N_0 is large enough. ■

Thus, it suffices to prove that both items of this proposition are satisfied. It is straightforward to check the first, much more difficult to check the second.

Lemma 4 *For each $z \in S^*$, the pair of mixed actions (α_z^0, β_z^0) is a Nash equilibrium of the matrix game with payoff $E(\gamma|z, \cdot, \cdot)$*

Proof. we first prove that α_z^0 is a best reply to β_z^0. For each ε, α_z^ε maximizes $E\left[\gamma^1(\alpha^\varepsilon, \beta^\varepsilon)|z, \cdot, \beta_z^\varepsilon\right]$. By letting ε go to zero, one gets that α_z^0 maximizes $E\left[\gamma^1|z, \cdot, \beta_z^0\right]$.

Observe that $\varepsilon \mapsto c(b; z, \alpha^\varepsilon, \beta^\varepsilon)$ is semi-algebraic, for every b, z, hence has a constant sign in a neighborhood of zero. Let $b \in B$. If it is the case that $c(b; z, \alpha^\varepsilon, \beta^\varepsilon) > 0$, for ε small enough, one has $\beta_z^\varepsilon(b) \leq \varepsilon$, by definition of Φ_2^ε. Therefore, for every action b in the support of $\beta_z^0$and ε small enough, one has $c(b; z, \alpha^\varepsilon, \beta^\varepsilon) = 0$ which means that b maximizes $E\left[\gamma^2(\alpha^\varepsilon, \beta^\varepsilon)|z, \alpha^\varepsilon, \cdot\right]$. One concludes as for player 1. ■

In the sequel, we give the main ideas of the proof that, for every h, the distribution Q^h of exit from D^h is controllable (for the continuation payoff γ). We let h be given and write D and Q for D^h and Q^h.

W.l.o.g, one may assume that, for every $(z, a, b) \in D \times A \times B$,

$$p(D|z, a, b) < 1 \Rightarrow p(D|z, a, b) = 0.$$

For the sake of the presentation, it is also convenient to assume that for any two distinct triples $(z_1, a_1, b_1), (z_2, a_2, b_2) \in D \times A \times B$ such that $p(D|z_1, a_1, b_1) = p(D|z_2, a_2, b_2) = 0$, the supports of the two distributions $p(\cdot|z_1, a_1, b_1)$ and $p(\cdot|z_2, a_2, b_2)$ are disjoint.

$>$From [6], we know that Q can be decomposed as a convex combination of unilateral exits and joint exits:

$$Q = \sum_{l \in L_1} \mu_l Q_l + \sum_{l \in L_2} \mu_l Q_l + \sum_{l \in L_3} \mu_l Q_l$$

where $Q_l = p(\cdot|z_l, a_l, \beta_{z_l}^0)$ (for some $(z_l, a_l) \in D \times A$) for $l \in L_1$, $Q_l = p(\cdot|z_l, \alpha_{z_l}^0, b_l)$ for $l \in L_2$, and $Q_l = p(\cdot|z_l, a_l, b_l)$ for $l \in L_3$. For $l \in L_3$, a_l and

b_l have the property $p(D|z_l, \alpha^0_{z_l}, b_l) = 1 = p(D|z_l, a_l, \beta^0_{z_l})$. We assume $\mu_l > 0$, for every l. Given our assumption on the supports, this decomposition is unique. To interpret (μ_l), denote by e the exit stage from D. For $l \in L_1$, μ_l is the limit as ε goes to zero of the probability that $(z^{e-1}, a^{e-1}) = (z_l, a_l)$. Similar interpretations are true for $l \in L_2, L_3$.

We refer to elements of L_1, L_2, L_3 as unilateral exits of player 1, unilateral exits of player 2, and joint exits.

To conclude this section, we explain what is the basic issue in proving that Q is controllable. As is shown in [6], this is straightforward *if* $Q_l\gamma^1 = \gamma^1(D)$ for every $l \in L_1$, and $Q_l\gamma^2 = \gamma^2(D)$ for every $l \in L_2$. Obtaining the first property is not difficult. However, there is no reason why the second property should hold. It might even be the case that $Q_l\gamma^2$ does depend on $l \in L_2$. In such a case, it is clear that player 2 would favor the unilateral exits $l \in L_2$ for which $Q_l\gamma^2$ is highest. It is also clear that no statistical test can be designed that would force player 2 to choose his various unilateral exits according to the weights μ_l, $l \in L_2$.

In the next section, we show on an example how the definition of the modified best-replies allows us to recover some properties of the quantities $Q_l\gamma^1$, for $l \in L_2$ (expected exit payoffs of player 1, associated to unilateral exits of player 2). We later sketch how to deal with the general case.

3.2 An example

Let us consider a game with two non-absorbing states, labeled z^1 and z^2. We shall not define the game completely. We rather assume that the basic data of the game (payoffs and transitions) are such that the Puiseux profile $(\alpha^\varepsilon, \beta^\varepsilon)$ has the following properties:

1. the unique maximal communicating set is $D = S^* = \{z^1, z^2\}$. In particular, the limit payoff $\gamma(z) = \lim \gamma(z, \alpha^\varepsilon, \beta^\varepsilon)$ is independent of the initial state z; we denote it $\gamma(D)$;

2. there exist $m, m' \in L_2$, such that $Q_m\gamma^2 < Q_{m'}\gamma^2 < \gamma^2(D)$.

Our goal is to deduce from the fixed-point properties of $(\alpha^\varepsilon, \beta^\varepsilon)$ a number of additional properties.

3.2.1 General remarks

We first derive elementary consequences of Lemma 4:

- for each $l \in L_1, Q_l \gamma^1 = \gamma^1(D)$;
- for each $l \in L_2, Q_l \gamma^2 \leq \gamma^2(D)$;
- Observe that $\gamma^2(z^1) = \max_B E\left[\gamma^2 | z^1, \alpha^0_{z^1}, \cdot\right]$, and that a similar relation holds for $\gamma^2(z^2)$.Since $\gamma^2(z^1) = \gamma^2(z^2)$, comparing the (limit) costs of two actions b and b' in the two states z and z' is equivalent to comparing $E\left[\gamma^2 | z, \alpha^0_z, b\right]$ and $E\left[\gamma^2 | z', \alpha^0_{z'}, b'\right]$:

$$\lim_{\varepsilon} c(b; z, \alpha^\varepsilon, \beta^\varepsilon) > \lim_{\varepsilon} c(b'; z', \alpha^\varepsilon, \beta^\varepsilon) \Leftrightarrow E\left[\gamma^2 | z, \alpha^0_z, b\right] < E\left[\gamma^2 | z', \alpha^0_{z'}, b'\right] \tag{3}$$

We now use the fact that $(\alpha^\varepsilon, \beta^\varepsilon)$ is a Puiseux profile. Thus, there are positive real numbers $p_z(b)$, and nonnegative numbers $d_z(b)$ $((z, b) \in S^* \times B)$, such $\beta^\varepsilon_z(b) \sim p_z(b)\varepsilon^{d_z(b)}$ as ε goes to zero. Similarly, $\alpha^\varepsilon_z(a) \sim p_z(a)\varepsilon^{d_z(a)}$. By definition of Φ_1, $p_z(a) > 0$ only if a maximizes $E\left[\gamma^1(\alpha^\varepsilon, \beta^\varepsilon) | z, \cdot, \beta^\varepsilon_z\right]$.

We conclude this section with a crucial observation. From (3) and the definition of Φ^ε_2, one deduces that, for any two pairs (z, b) and (z', b'),

$$E\left[\gamma^2 | z, \alpha^0_z, b\right] < E\left[\gamma^2 | z', \alpha^0_{z'}, b'\right] \Rightarrow d_z(b) > 2d_{z'}(b'). \tag{4}$$

3.2.2 D-graphs and degrees of transitions

Recall that D is a communicating set. From Freidlin-Wentzell's formula, we know that the exit distribution Q from D can be expressed in terms of D-graphs. We shall here have a closer look.

Since D is a communicating set, there is some pair (a, b) such that $p(D | z^1, a, b) = 1$ and $p(z^2 | z^1, a, b) > 0$. Define the degree of the transition from z^1 to z^2, $d(z^1 \to z^2)$ as the minimum of $d_{z^1}(a) + d_{z^1}(b)$ over such pairs (a, b). Define $d(z^2 \to z^1)$ similarly.

For $l \in L_1 \cup L_2 \cup L_3$, we define the degree $\deg(l)$ of the exit labeled l as follows. If $l \in L_3$ with $Q_l = p(\cdot | z^2, a_l, b_l)$, we set

$$\deg(l) = d(z^1 \to z^2) + d_{z^2}(a_l) + d_{z^2}(b_l).$$

If $l \in L_1$ with $Q_l = p(\cdot|z^2, a_l, \beta_{z^2})$, we set

$$\deg(l) = d(z^1 \to z^2) + d_{z^2}(a_l).$$

The degree of other types of exits is defined accordingly. The following observation is an immediate consequence of Freidlin-Wentzell's formula.

Lemma 5 $\deg(l)$ *is independent of* $l \in L_1 \cup L_2 \cup L_3$.

3.2.3 Exits of player 2 and continuation payoffs of player 1

We derive some implications of Lemma 5 and (4). Since $Q_m\gamma^2 < Q_{m'}\gamma^2$,

$$d_{z_m}(b_m) > d_{z_{m'}}(b_{m'}). \tag{5}$$

Since $\deg(m) = \deg(m')$, it must be the case that $z_m \neq z_{m'}$. To fix the ideas, we assume $z_m = z^1$, and $z_{m'} = z^2$. For similar reasons, if $p(S^*|z^1, \alpha^0_{z^1}, b) < 1$, then one has $E\left[\gamma^2|z^1, \alpha^0_{z^1}, b\right] \leq 1$. Another consequence is that, for any $l \in L_2$, $Q_l\gamma^2 = Q_m\gamma^2$ if $z_l = z^1$, and $Q_l\gamma^2 = Q_{m'}\gamma^Z$ if $z_l = z^2$. We divide the unilateral exits L_2 of player 2 into L_2^1 and L_2^2 accordingly.

Denote by $\overline{Q}_1$ and $\overline{Q}_2$ the renormalizations of Q over exits in L_2^1 and L_2^2 respectively:

$$\overline{Q}_i = \frac{\sum_{l \in L_2^i} \mu_l Q_l}{\sum_{l \in L_2^i} \mu_l}, .i = 1, 2.$$

We shall prove that $\overline{Q}_1\gamma^1 = \gamma^1(D) = \overline{Q}_2\gamma^1$. In words, this means that player 1 is indifferent between the two classes of unilateral exits of player 2.

Lemma 6 $\overline{Q}_1\gamma^1 = \gamma^1(D)$.

Proof. We show that player 1 is able to *block* the transition from z^1 to z^2. By this, we mean that, in order to reach z^2 from z^1without leaving $\{z^1, z^2\}$, it is necessary that player 1 perturbs $\alpha^0_{z^1}$. We argue by contradiction. Assume that, for some b, $p(S^*|z^1, \alpha^0_{z^1}, b) = 1$ and $p(z^2|z^1, \alpha^0_{z^1}, b) > 0$. Observe that $E\left[\gamma^2|z^1, \alpha^0_{z^1}, b\right] = \gamma^2(D)$. Therefore,

$$d_{z^1}(b_m) > 2d_{z^2}(b_{m'}) \text{ and } d_{z^1}(b_m) > 2d_{z^1}(b).$$

Since $d(z^1 \to z^2) \leq d_{z^1}(b)$, and $d(z^2 \to z^1) \geq 0$, this yields

$$d(z^2 \to z^1) + d_{z^1}(b_m) > d(z^1 \to z^2) + d_{z^2}(b_{m'})$$

which contradicts $\deg(m) = \deg(m')$.

By definition of Φ_1, α^0 is a best reply to β^ε, for every ε small enough. In particular,

$$\lim_\varepsilon \gamma^1(z^1, \alpha^0, \beta^\varepsilon) = \lim_\varepsilon \gamma^1(z^1, \alpha^\varepsilon, \beta^\varepsilon) = 1.$$

Starting from z^1, the play cannot reach z^2 under $(\alpha^0, \beta^\varepsilon)$ (in particular, $\{z^1\}$ is an ergodic set under (α^0, β^0)). It will eventually reach an absorbing state, when player 2 plays a unilateral exit. It is easy to check that

$$\lim_\varepsilon \gamma^1(z^1, \alpha^0, \beta^\varepsilon) = \overline{Q}_1 \gamma^1.$$

■

Lemma 7 $\overline{Q}_2 \gamma^1 = \gamma^1(D)$

Proof. the proof is more subtle than the previous one. The previous argument relies on the fact that, when player 1 plays α^0, there is no way for player 2 to reach z^2, starting from z^1. When one exchanges z^1 and z^2, the corresponding fact needs not hold. It might for instance be the case that z^2 is transient under (α^0, β^0). It might thus be the case that

$$\lim_\varepsilon \gamma^1(\mathsf{z}^2, \alpha^0, \beta^\varepsilon) = \overline{Q}_1 \gamma^1$$

therefore repeating the previous argument starting from z^2 gives nothing new.

However, the previous argument works in the case where player 1 can block the transition from z^2 to z^1: there is no $b \in B$, such that $p(\{z^1, z^2\} | z^2, \alpha^0_{z^2}, b) = 1$ and $p(z^1 | z^2, \alpha^0_{z^2}, b) > 0$. Thus we may assume that such a b does exist. Denote it b^*.

We shall infer additional properties by modifying the *degrees* of the actions of player 1. Assume that, for every pair such that $d_z(a) > 0$ the degree of a is modified to $\overline{d}_z(a)$ (we do not rule out the case $\overline{d}_z(a) = d_z(a)$), and denote by $\overline{\alpha}^\varepsilon$ the resulting strategy. Formally,

$$\overline{\alpha}^\varepsilon_z(a) = \frac{p_z(a)\varepsilon^{\overline{d}_z(a)}}{\sum_A p_z(a')\varepsilon^{\overline{d}_z(a')}}$$

where we set for simplicity $\overline{d}_z(a) = 0$ if $d_z(a) = 0$. Clearly, $\overline{\alpha}^\varepsilon$ belongs to $\Phi_1(\beta^\varepsilon)$ and thus

$$\lim_\varepsilon \gamma^1(\overline{\alpha}^\varepsilon, \beta^\varepsilon) = \gamma^1.$$

It is also clear that $\lim_\varepsilon \gamma^1(\overline{\alpha}^\varepsilon, \beta^\varepsilon) = \sum_{l \in L_1} \overline{\mu}_l Q_l + \sum_{l \in L_2} \overline{\mu}_l Q_l + \sum_{l \in L_3} \overline{\mu}_l Q_l$ for some weights $(\overline{\mu}_l)_l$ (some of these weights might here be zero).[2]

We now show that, by a proper choice of the degrees $\overline{d}_z(a)$, for $(z, a) \in S^* \times A$, the induced weights $\overline{\mu}_l$ will satisfy

C.1 $\overline{\mu}_l = 0$, for each $l \in L_1 \cup L_3$

C.2 $\overline{\mu}_l = \frac{\mu_l}{\sum_{L_2} \mu_k}$, for $l \in L_2$.

For $l \in L_1 \cup L_3$, we fix $\overline{d}_{z_l}(a_l) > d_{z_l}(a_l)$: we increase the degree of all the actions of player 1 which are involved in unilateral exits of player 1 or in joint exits. Clearly, the new degree $\overline{\deg}(l)$ of any $l \in L_1 \cup L_3$ is higher than $\deg(l)$.

We shall now prove that $\overline{\deg}(l) = \deg(l)$, for any $l \in L_2$. This will provide **C.1**. Since degrees of actions of player 2 are unchanged, we need to prove that

$$\overline{\deg}(z^1 \to z^2) = \deg(z^1 \to z^2)$$
$$\overline{\deg}(z^2 \to z^1) = \deg(z^2 \to z^1)$$

To get **C.2,** one needs more. Let $(a, b) \in A \times B$ be any pair of actions which is involved in the transition $z^1 \to z^2$:

$$p(\{z^1, z^2\}|z^1, a, b) = 1, \; p(z^2|z^1, a, b) > 0 \text{ and } d_{z^1}(a) + d_{z^1}(b) = d(z^1 \to z^2).$$

We shall prove that $\overline{d}_{z^1}(a) = d_{z^1}(a)$. Results regarding transitions from z^2 to z^1 can be obtained by a similar proof.

Assume that $\overline{d}_{z^1}(a) > d_{z^1}(a)$. By definition of the new degrees, this means that $a = a_l$, for some $l \in L_1 \cup L_3$. Clearly, $l \in L_1$ would contradict the fact that $p(\{z^1, z^2\}|z^1, a, \beta^0_{z^1}) = 1$. Thus, $l \in L_3$: one has $Q_l = p(\cdot|z^1, a, b_l)$, for some $b_l \in B$ such that $p(\{z^1, z^2\}|z^1, \alpha^0_{z^1}, b_l) = 1$.

By Lemma 5, one has $\deg(m') = \deg(l)$, which reads

$$\deg(z^1 \to z^2) + d_{z^2}(b_{m'}) = \deg(z^2 \to z^1) + d_{z^1}(a) + d_{z^1}(b_l). \tag{6}$$

Recall now that $d(z^1 \to z^2) = d_{z^1}(a) + d_{z^1}(b)$ and that, by definition of b^*, one has $d(z^2 \to z^1) \leq d_{z^2}(b^*)$. Substituting in (6) yields

$$d_{z^2}(b_{m'}) \leq d_{z^1}(b_l) + d_{z^2}(b^*) - d_{z^1}(b) \leq d_{z^1}(b_l) + d_{z^2}(b^*). \tag{7}$$

Observe now that neither b_l in state z^1, nor b^* in state z^2 is a unilateral exit of player 2: one has $p(\gamma^2|z^1, \alpha^0_{z^1}, b_l) = \gamma^2(D) = p(\gamma^2|z^2, \alpha^0_{z^2}, b^*)$. By (4), this implies

$$d_{z^2}(b_{m'}) > 2d_{z^1}(b_l) \text{ and } d_{z^2}(b_{m'}) > 2d_{z^2}(b^*),$$

[2] It is not clear that the decomposition of the new exit distribution will involve only the exits labeled $l \in L$. This will however be the case, provided the degrees are only slightly modified.

which is in contradiction with (7). It is now not difficult to get the result. ■

What have we proven so far ? Let us rephrase our results:

- denote by $\overline{L}_2$ the set of $l \in L_2$ which satisfy $Q_l\gamma^2 < \gamma^2(D)$; there is a partition (L_2^1, L_2^2) of $\overline{L}_2$ into level sets for $Q_l\gamma^2$;
- to each set L_2^i $(i = 1, 2)$ is associated a communicating subset $F^i$of D : $F^1 = \{z^1\}$, and $F^2 = \{z^2\}$ or $\{z^1, z^2\}$ depending on whether or not player 1 can block the transition from z^2 to z^1. For each i, player 1 can block the transition from F^i to the remaining part of D (if $F^i = D$, this is an empty statement). Moreover,
  - the expectation of γ^1 under the renormalized exit distributions $Q_{L_2^1}$ and $Q_{L_2^2}$ is $\gamma^1(D)$.
  - in F^i, there is no unilateral exit of player 2 which is better than exits in L_2^i :
  $$E\left[\gamma^2|z_l, \alpha^0_{z_l}, b_l\right] \geq E\left[\gamma^2|z, \alpha^0_z, b\right],$$
  for every $l \in L_2^i$, $(z, b) \in F^i \times B$ such that $p(\{z^1, z^2\}|z, \alpha^0_z, b) < 1$.
- finally, $Q_l\gamma^1 = \gamma^1(D)$ for every $l \in L_1$, $E\left[\gamma^1|z, a, \beta^0_z\right] \leq \gamma^1(D)$ and $E\left[\gamma^2|z, \alpha^0_z, b\right] \leq \gamma^2(D)$, for every $(z, a, b) \in D \times A \times B$.

As shown in [6], this is enough to ensure that Q is controllable. The corresponding profile uses public lotteries, performed by player 1.

3.3 The general case

We briefly indicate how to generalize the previous example. We make no attempt at a proof.

As above, let D be a maximal communicating subset, and write the decomposition of the corresponding exit distribution as

$$Q = \sum_{l \in L_1} \mu_l Q_l + \sum_{l \in L_2} \mu_l Q_l + \sum_{l \in L_3} \mu_l Q_l$$

The problem is to find a partition $(L^1, ..., L^H)$ of $\overline{L}_2 = \{l \in L_2, Q_l\gamma^2 < \gamma^2(D)\}$ and communicating subsets $(F^1, ..., F^H)$ of D that have the properties of the previous section.

Let $l \in \overline{L}_2$. Denote by $D^1 \subset D^2 ... \subset D^M$ the communicating subsets of D which contain z_l. Denote by $D(l)$ the first one in this sequence that has the property that it is much more difficult to leave $D(l)$ than to reach z_l starting from $D(l)$. We will not define this property formally. It is an extension of the property that we used in the case of two non-absorbing states for F^i, namely that player 1 can block the transition from F^i to the remaining states of D.

The sets $L^1, ..., L^H$ and $F^1, ..., F^H$ are obtained as follows. Denote first by $\overline{L}^1, ..., \overline{L}^P$ the partition of $\overline{L}_2$ into level sets for $l \mapsto Q_l \gamma^2$. For each p, we look at the equivalence classes of the relation $\mathcal{R}_p$ defined on $\overline{L}^p$ by

$$l \mathcal{R}_p l' \Leftrightarrow D(l) = D(l')$$

The sets $L^1, ..., L^H$ are all the equivalence classes of all the relations $\mathcal{R}_p$, $1 \leq p \leq P$. For $1 \leq h \leq H$, we set $D^h = D(l)$, where $l \in L^h$.

References

[1] Blackwell, D. (1962). Discrete dynamic programming. *Annals of Mathematical Statistics, 331:719-726.*

[2] Everett, H. (1957). Recursive games, in *Contributions to the Theory of Games,* Annals of Mathematical Studies, 39, Princeton University Press.

[3] Flesch, J., F. Thuijsman and O.K. Vrieze (1996). Recursive repeated games with absorbing states, *Mathematics of Operations Research,* 21:1016-1022.

[4] Neyman, A. (1999). *this book.*

[5] Solan, E. (1997). Stochastic games with two non-absorbing states, Discussion Paper 160, *Center for Rationality and Interactive Decision Theory.*

[6] Solan, E. (1999). General tools- Perturbations of Markov chains, *this book.*

[7] Vieille, N. (1992). Thèse de doctorat, Université Paris VI.

[8] Vieille, N. (1997). Two-person stochastic games II: the case of recursive games, *Cahiers du CEREMADE 9747.*